

# ネットワーク構造の統計的な推定手法について

増田 直紀<sup>1,2</sup>

(受付 2011 年 12 月 2 日 ; 改訂 2012 年 2 月 3 日 ; 採択 3 月 1 日)

## 要 旨

ここ 15 年ほど、ネットワークの研究が盛んになり、ネットワーク科学、複雑ネットワークといった研究分野で呼ばれるようになってきている。ここで言うネットワークとは、グラフ理論の指すグラフと同義である。様々な種類の実際のグラフがデータとして入手できるようになってきたことなどを背景として、統計物理学、応用数学、ウェブ工学などの分野の研究者が主になって研究が始まり、現在は様々な研究領域を巻き込んで研究が拡大している。データを相手とする以上、統計科学は様々な形でこれらのネットワーク研究に応用される。本稿では、ネットワークの構造をデータから最尤推定する 2 つの方法について簡単に紹介する。1 つ目は Newman と Leicht が提案したネットワーク構造の最尤推定手法である。この手法では、1 つのグラフがいくつかのグループに分割され、同じグループに属する頂点は統計的な意味で同じような所へ直接リンクしやすいと仮定した上で、結合確率や頂点のグループ分けを規定する変数の値を最尤推定する。2 つ目は、最大エントロピー法による相互作用の推定手法である。これは古典的な方法ではあるが、近年、多点で同時記録される神経活動データ解析への応用とそれにまつわる理論の発展が進んでいる。

キーワード：グラフ、ネットワーク、コミュニティ構造、EM アルゴリズム、イジングモデル。

## 1. はじめに

本稿で扱う「ネットワーク」は、グラフ理論の指す「グラフ」と同値である。すなわち、1 つのネットワークは、図 1 に例を示すように、頂点の集合と枝の集合からなる。枝は 2 つの頂点を接続し、有向である場合と無向である場合がある。ネットワーク(の)科学、あるいは、複雑ネットワークとは、ネットワークの諸側面を扱う学際的な研究分野であると言える。大量の実データが整備され、統計物理学の手法も駆使されて、1998 年頃から発展してきた。ネットワーク研究一般については、日本語でも増田・今野(2006, 2010)、増田・中丸(2006)、増田(2008a, 2008b, 2010)、青山 他(2008)等の様々な紹介があるので(増田・今野, 2010 の巻末には、解説付きで詳しい文献リストがある)、本稿では詳しく触れない。

簡単に述べると、三角形が多い(クラスター性)、頂点間の平均距離が短い(スモールワールド性。三角形が多いことと合わせた上でスモールワールド性ということもある)、よくつながった頂点とあまりつながっていない頂点が共存している(スケールフリー性)、といった構造的な特徴が、色々なネットワークの実例や数理モデルの解析を通じて明らかにされた。また、ネッ

<sup>1</sup> 東京大学大学院 情報理工学系研究科：〒113-8656 東京都文京区本郷 7-3-1

<sup>2</sup> 科学技術振興機構さきがけ：〒332-0012 埼玉県川口市本町 4-1-8

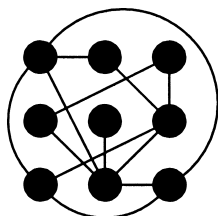


図 1. ネットワークの例.

ネットワークの故障耐性, その上での感染症ダイナミクス, 同期ダイナミクス, などの現象モデルについての知見が急速に積み立てられてきたことも特徴である.

進化, 多様性, 生物といった本特集のテーマを踏まえてネットワーク関連の総説を執筆する最も実直な道筋は, 直接この話題に触れる内容を概説することであろう. 例えば, ネットワーク上の進化という研究課題がある (Lieberman et al., 2005; Antal et al., 2006; Sood et al., 2008; Masuda and Ohtsuki, 2009). 最も単純なそのようなモデルの例では, ネットワーク ( $\equiv$  グラフ) 上の各頂点が任意の時刻で黒か白のどちらかの状態をとる. そして, 各頂点は, 自分と異なる色の隣接頂点数に比例するレートを持つポアソン過程に従うイベントの時刻で, 自分の色を反転する. 周りの色に同調するのである. これは与えられたネットワーク上のマルコフ過程である. ネットワークが有限ならば, いずれ, 全ての頂点が黒または白になって, ダイナミクスは終了する. このモデルは, 確率論や統計物理学では投票者モデルと呼ばれる (Liggett, 1985; Durrett and Levin, 1994; Redner, 2001; Castellano et al., 2009). また, 集団遺伝学における, 中立な変異体が侵入したときの固定確率 (ある黒と白の配置から始まって, 全部が例えば黒になってダイナミクスが終了する確率) の問題と同じである (Ewens, 2010). 状態遷移規則がより複雑である「ネットワーク上の進化ゲーム」も数理生物学, 統計物理学, その他の社会シミュレーションを扱う研究分野などで研究されている (Ohtsuki et al., 2006; Szabó and Fáth, 2007; 増田, 2008c; 増田・今野, 2010 の 9.1 節). 自分が黒と白のどちらであるかに応じて色を反転するレートが異なり, さらに, 反転レートが隣接頂点の色だけでなくそのさらに隣の頂点の色にも依存する, という意味で複雑である.

しかし, 統計学とネットワーク研究の接点について個人的に考えた結果として, 本稿でそのような話題を概説することはしない. ネットワーク上の進化ダイナミクスは, 通常は確率モデルであるが, モデルの変数値をデータから推定することはあまり期待されていないように思われる. 実際には構造が部分的に未知でありうることも考え合わせると, 現実のネットワークは複雑である. さらに, その上で何らかの進化ダイナミクスが進行しているとする. さらに, ネットワークの非一様性のために, 各頂点は, モデルが想定している以外の意味でも異なるダイナミクスをもつかもしい. したがって, 得られる実測データからモデルを推定することは, 変数が多すぎて難しいかもしれない. そもそも, ネットワーク上の進化ダイナミクスについての実測データは少ない. 今後もしばらくそうであると思われる.

それよりも, ネットワークの不完全な観測データに基づいて, ネットワーク構造を推定する, という問題が, 統計学の応用が期待される喫緊の課題であると思われる. ネットワーク科学の研究全体も, モデルを提案して解析して理論や個体ベースの数値計算 (例えば進化ダイナミクスの進行を直接数値計算すること) で閉じた研究を行うことは一段落して, 大量のデータをどう扱うか, 不完全なデータから何を結論するか, という研究に重心が移ってきているように感じる. ネットワークという研究分野がこれから 10~20 年の範囲で真に実用的な価値を出した

めにも、実データと向き合うことは避けて通れない。

そこで、本稿では、「多様性」を、ネットワークの複雑性やその上に展開する複雑な(実際には、頂点によって異なりうるという意味で複雑な)過程として広く解釈し、ネットワークの構造推定に関する2つの統計的な方法について概説する。1つ目は、静的なネットワーク構造の統計モデルである。2つ目は、ネットワーク上の(進化とは異なる)ある種のダイナミクスの観測データから、(しかし、ダイナミクスの情報は切り捨てて)ネットワークの構造を推定する問題設定である。

## 2. グループ構造を仮定した、ネットワーク構造の統計モデル

枝の存在ないし非存在は、そもそも確率的な事象であると考えるのは現実的である。これと関係して、ネットワークを生成するモデルのほとんどは確率モデルである(増田・今野, 2010)。この場合、現実のネットワークは、生成モデルの確率空間(以下では、統計モデルと呼ぶ)からサイコロをふって得られた実現値の1つということになる。どのような統計モデルが適切であろうか。

あるいは、逆に、全ての各頂点对について、何らかの手段で枝の有無が確定的に(背後にある統計モデルの1実現値として得られたというわけではなく)わかっているとしよう。この場合でも、頂点数が多いと、ネットワークを可視化したり隣接行列を眺めたりすることだけからネットワークについて有用な情報を取り出すことは困難である。例えば、頂点が  $N = 10^4$  個あれば、ネットワークが疎(各頂点の次数、すなわち、枝の数が  $o(N)$  であること)だとしても、可視化の結果は枝が多すぎて解釈しにくい。実際に我々が見ているネットワークは隣接行列の要素数や枝数ほどには独立かつ重要な要素がないならば、ネットワークのより簡約な表現を作ることができるかもしれない。

そこで、観測された1つのネットワークはある統計モデルから生成された1サンプルであると仮定して統計モデルを規定する変数を最尤推定する、という一般的な道筋が考えられる。統計モデルの変数の数は隣接行列の要素数よりは遙かに少なくなるようにする。本節では、Newman and Leicht (2007) によるこのような手法の1つを紹介する(Leicht et al., 2007 も参考)。

頂点数を  $N$  で表す。隣接行列を  $A$  とする。すなわち、頂点  $i$  から  $j$  に枝があるときには  $A_{ij} = 1$ 、ないときは  $A_{ij} = 0$  である。本節では有向グラフを扱う。すなわち、 $A_{ij}$  と  $A_{ji}$  は一般的には等しくないとする。 $A$  は観測される量である。

各頂点は  $c$  個あるグループのうちの1つに属すると仮定する。頂点  $i$  が属するグループの番号を  $g_i$  ( $1 \leq g_i \leq c$ ,  $1 \leq i \leq N$ ) とする。これは大きな仮定だが、現実に見られる多くのネットワークがそのようなグループ構造を持っていることが以前の様々な研究で示されている、という背景がある。人間の社会ネットワークがその好例であるが、他の種類のネットワークも往々にしてそうである。この前提知識を利用して、ネットワークの簡約な表現をつくるのである。

グループの定義として、同じグループに属する頂点は似たつながり方をすると仮定する。特に、通常は、同じグループ内では結合が密で、異なるグループ間では結合が疎であることが仮定される。そのようなネットワークの構造は、コミュニティ構造、モジュール構造などと呼ばれる。多くのネットワークがコミュニティ構造をもつのである。同じコミュニティに属する頂点どうしは、機能や所持している情報の意味でも近いことが経験上多い。

与えられたネットワークをコミュニティに分割するアルゴリズムの研究は盛んであり、増田・今野(2010)の2.6節や Fortunato (2010) が総合的な解説として参考になる。分割の良し悪しを決める基準として、Newman and Girvan (2004) で導入されたモジュラリティという指標がここ7, 8年で有名になったが、この指標を最大化するのは「1つの」指針であるととらえるのが

妥当である．特に，有向グラフのコミュニティ分割の手法については，十分な決着がついていないようである．また，分割に要する計算量も，コミュニティ分割アルゴリズムの設計に際して気にされる基準の 1 つである．ただし，本節で説明する Newman and Leicht (2007) の手法は，コミュニティ構造，すなわち，内部では結合が密で外部とは結合が疎であることを仮定せずとも適用できる．同じグループ内では結合が疎で，異なるグループ間では結合が密である（二部グラフを想像されたい）ようなネットワーク構造の抽出がその例である．

このようにグループがネットワーク構造の単位であるというとらえ方をすると，1 つのネットワーク，すなわち，1 回の観測結果が，あたかも，繰り返し観測の結果であるように解釈できる．すなわち，各コミュニティに属する頂点は通常多数あるので，例えば，コミュニティ 1 とコミュニティ 2 を結ぶ枝の有無についての観測が，単一のネットワークから多数回得られる．したがって，統計的な手法を用いやすい．

グループ  $r$  に属する 1 つの頂点から出る 1 本の枝を考える． $\theta_{ri}$  を，この枝が頂点  $i$  に接続する確率とする．また， $\pi_r$  を，各頂点がグループ  $r$  に属する確率とする．これらの変数は，非観測データ  $g_i$  とともに未知であり，観測された隣接行列  $A$  から推定されるべきものである．正規化条件は

$$(2.1) \quad \sum_{r=1}^c \pi_r = 1, \quad \sum_{i=1}^N \theta_{ri} = 1$$

である．モデル上は，自己ループは許容される．

以下， $\pi = \{\pi_r\}$  と  $\theta = \{\theta_{ri}\}$  の最尤推定法を説明する．データは，観測されるデータ（観測データ） $A$  と観測されない潜在データ（潜在変数） $\{g_i\}$  であると考える． $g = \{g_i\}$  とし，尤度を

$$(2.2) \quad \Pr(A, g | \pi, \theta) = \Pr(A | g, \pi, \theta) \Pr(g | \pi, \theta)$$

と書く． $\Pr(\cdot | \cdot)$  は条件付き確率を表す． $A_{ij} \in \{0, 1\}$  に注意すると

$$(2.3) \quad \Pr(A | g, \pi, \theta) = \prod_{i=1}^N \prod_{j=1}^N (\theta_{g_i, j})^{A_{ij}},$$

$$(2.4) \quad \Pr(g | \pi, \theta) = \prod_{i=1}^N \pi_{g_i},$$

となる．式 (2.3) で， $A_{ij} = 0$  のときに  $(1 - \theta_{g_i, j})$  といった項が現れるわけではないことに注意する． $\theta_{g_i, j}$  は，頂点  $i$  と頂点  $j$  が隣接する確率ではなく，頂点  $i$  から出る 1 本の枝があるという条件をつけたもとでの， $i$  と  $j$  が隣接する確率だからである．すると，観測データ  $A$  は，隣接行列そのものではなく，各頂点からの出次数を固定したもとでの隣接行列，ということになる．原論文 (Newman and Leicht, 2007) はこのことを明確に説明していない．以降，原論文が基づくと思われる，出次数は固定という解釈に立脚して話を進める．式 (2.2), (2.3), (2.4) より，完全対数尤度は

$$(2.5) \quad \ln \Pr(A, g | \pi, \theta) = \sum_{i=1}^N \left[ \ln \pi_{g_i} + \sum_{j=1}^N A_{ij} \ln \theta_{g_i, j} \right]$$

となる．

データの一部である  $g$  が未知なので，EM アルゴリズムを用いる．すなわち，周辺尤度  $\Pr(A | \pi, \theta)$  の代わりに潜在変数  $g$  の事後分布についての完全対数尤度の期待値

$$(2.6) \quad \sum_{g_1=1}^c \cdots \sum_{g_N=1}^c \Pr(g | A, \pi, \theta) \sum_{i=1}^N \left[ \ln \pi_{g_i} + \sum_{j=1}^N A_{ij} \ln \theta_{g_i, j} \right]$$

$$= \sum_{i=1}^N \sum_{r=1}^c \Pr(g_i = r | A, \pi, \theta) \left[ \ln \pi_r + \sum_{j=1}^N A_{ij} \ln \theta_{rj} \right]$$

を考える． $\Pr(g_i = r | A, \pi, \theta)$  を計算するには  $\pi$  と  $\theta$  が必要である．一方，最尤法で  $\pi$  と  $\theta$  を計算するには  $\Pr(g_i = r | A, \pi, \theta)$  が必要である．そこで，EM アルゴリズムである．

$\Pr(g_i = r | A, \pi, \theta)$  は， $\pi$  と  $\theta$  から，EM アルゴリズムの E ステップとして以下のように求める：

$$\begin{aligned} (2.7) \quad \Pr(g_i = r | A, \pi, \theta) &= \frac{\Pr(A, g_i = r | \pi, \theta)}{\Pr(A | \pi, \theta)} \\ &= \frac{\left[ \pi_r \prod_{j=1}^N (\theta_{rj})^{A_{ij}} \right] \left[ \prod_{1 \leq \ell \leq N, \ell \neq i} \sum_{r'=1}^c \pi_{r'} \prod_{j=1}^N (\theta_{r'j})^{A_{\ell j}} \right]}{\prod_{\ell=1}^N \sum_{r'=1}^c \pi_{r'} \prod_{j=1}^N (\theta_{r'j})^{A_{\ell j}}} \\ &= \frac{\pi_r \prod_{j=1}^N (\theta_{rj})^{A_{ij}}}{\sum_{r'=1}^c \pi_{r'} \prod_{j=1}^N (\theta_{r'j})^{A_{ij}}} . \end{aligned}$$

式 (2.7) で  $\Pr(g_i = r | A, \pi, \theta)$  が得られたら，対数尤度 (2.6) を最大化する  $\pi, \theta$  を求めればよい．EM アルゴリズムの M ステップである．具体的には，正規化条件 (2.1) に対応するラグランジュ未定定数を  $\lambda_1, \lambda_2$  として

$$(2.8) \quad [\text{式 (2.6)}] + \lambda_1 \left( \sum_{r=1}^c \pi_r - 1 \right) + \lambda_2 \left( \sum_{j=1}^N \theta_{rj} - 1 \right)$$

を考える．式 (2.8) を  $\pi_r$  ( $1 \leq r \leq c$ ) で偏微分し， $\lambda_1$  を消去し， $\sum_{r=1}^c \pi_r = 1$  が満たされるように比例係数を選ぶと

$$(2.9) \quad \pi_r = \frac{1}{N} \sum_{i=1}^N \Pr(g_i = r | A, \pi, \theta)$$

となる．同様に，式 (2.8) を  $\theta_{rj}$  ( $1 \leq r \leq c, 1 \leq j \leq N$ ) で偏微分し， $\lambda_2$  を消去し， $\sum_{j=1}^N \theta_{rj} = 1$  が満たされるように比例係数を選ぶと

$$(2.10) \quad \theta_{rj} = \frac{\sum_{i=1}^N A_{ij} \Pr(g_i = r | A, \pi, \theta)}{\sum_{i=1}^N k_i^{\text{out}} \Pr(g_i = r | A, \pi, \theta)}$$

となる．ただし， $k_i^{\text{out}} \equiv \sum_{\ell=1}^N A_{i\ell}$  は頂点  $i$  の出次数である．

$A = (A_{ij})$  が与えられているもとの式 (2.7), (2.9), (2.10) を反復すると， $\pi, \theta, \Pr(g_i = r | A, \pi, \theta)$  を求めることができる．頂点の(確率的な)グループ分けの情報は  $\Pr(g_i = r | A, \pi, \theta)$  に含まれている． $\theta$  は得られたグループ分けの内容を与えている．すなわち， $\theta_{rj}$  は任意のグループ  $r$  と  $r'$  (正確には，グループ  $r'$  内の頂点  $j$ ) の結びつきやすさを表す．

Newman and Leicht (2007) では，無向グラフの場合 ( $A_{ij} = A_{ji}$ ) についても同様の手法をつくらることができることが説明されている．

なお，論文 Newman and Leicht (2007) のウェブサイト<sup>1)</sup>に，このアルゴリズムのプログラムコードが公開されている．Newman のウェブサイト<sup>2)</sup>には，この方法だけでなく，他のコミュニティ分割手法やネットワークの最尤推定手法のコードも公開されている．コードが公開されている関連手法の例として，ネットワークが，階層的な統計モデル(観測されない)から生成されていると仮定する方法が Clauset et al. (2008) で提案されている．すなわち，各頂点は，ある系統樹の末端の葉であるとする．2 頂点が隣接する確率は，2 頂点に対応する葉が系統樹の根元に向かって上っていくときにどの高さで出会うか，で規定されるとする．この確率と系統樹

の構造がモデルである。系統樹が与えられれば、観測されたネットワークの尤度は簡単に計算できる。そこで、系統樹が与えられたもとでのパラメータ値を最尤推定で求め、系統樹の構造はマルコフ連鎖モンテカルロ法で求める。

### 3. 最大エントロピー法を用いたネットワーク構造の推定

ネットワーク構造はわからないが、各頂点の活動は計測できる場合がしばしばある。脳のニューロン活動はその例である。多点電極などを用いれば、複数のニューロン活動を同時に、高い時間解像度で、計測できる。多点電極の場合、1つの電極は1つのニューロンに対応するわけではなく、周囲にある複数のニューロンの活動が重なった信号が得られ、マルチユニット活動(MUA)と呼ばれる。そのような活動の時系列を多点について計測でき、観測が行われた複数の点から成るネットワークは、ニューラル・ネットワークについてかなりの知識を与えてくれる。

神経系の実験において、結合の有無を決めることは、活動の有無を計測することよりもはるかに難しい。解剖学的、組織学的な方法などで結合の有無を決めたりするが、多大な労力を要する。

このような背景もあって、多点の神経活動から結合の度合いを推測する手法がいくつか考えられている。ここでは、そのような手法のうち、Schneidman et al. (2003, 2006), Shlens et al. (2006), Tang et al. (2008), Yeh et al. (2010)などで用いられているエントロピー最大化原理を用いる方法をごく簡単に紹介する。他の関連する解説に例えば田中(2006)の5.2節近辺がある。エントロピー最大化原理は統計物理学で Jaynes (1957)によってなされた定式化であり、以前から他の分野では(ネットワークという言い方はあまりせずに)使われているようである。近年は、対数線形モデル、情報幾何の神経活動解析への応用という文脈でも研究が行われている。ごく最近の解説に島崎(2011)がある。

頂点数  $N$  の、ある未知のネットワークが与えられているとする。各頂点の活動がいくつかの時刻に渡って観測されているとする。離散時刻を仮定し、各離散時刻  $1 \leq t \leq T$  で、各頂点  $i$  ( $1 \leq i \leq N$ ) は活動( $\sigma_i(t) = 1$ )か非活動( $\sigma_i(t) = -1$ )のどちらかの状態をとるとする。そして、各頂点の活動量の時間平均

$$(3.1) \quad \langle \sigma_i \rangle \equiv \frac{1}{T} \sum_{t=1}^T \sigma_i(t) \quad (1 \leq i \leq N)$$

と各頂点对の時間平均の意味での活動相関

$$(3.2) \quad \langle \sigma_i \sigma_j \rangle \equiv \frac{1}{T} \sum_{t=1}^T \sigma_i(t) \sigma_j(t) \quad (1 \leq i \leq j \leq N)$$

が観測されているとする。

こう書いた時点で、暗黙に仮定している事項がいくつかある。まず、時間平均をとった量のみを観測量とするので、時間相関を無視している。実際のデータは連続時間で観測されることも多い。それを離散化幅  $\Delta t$  で離散化することにより、ここで仮定したような離散時刻のデータが得られる。 $\Delta t$  を大きくするほど、時間相関を無視するという仮定は正当化されやすい。しかし、 $\Delta t$  が大きいと、サンプル数  $T$  が減るのでモデルの推定が危うくなりうる。なお、神経系のデータを扱うには、 $\Delta t \approx 20$  ms がよく用いられている。また、全ての時刻のデータを同等に扱うので、データについて、時間的に無相関であるだけでなく定常性があることも暗黙に仮定されている。時間相関や因果性がある場合については後述する。

さて、式 (3.1), (3.2) の制約のもとで、エントロピーを最大化する分布  $P_2(\sigma_1, \dots, \sigma_N)$  を求

めよう． $P_2$  は  $2^N$  個の点の上にマスを持つ離散確率分布である．すなわち， $2^N$  種類のネットワークの状態がありうる．与えられた制約のもとで最もランダムな分布を求めるのである．

最大化すべき量は

$$(3.3) \quad \sum_{\sigma_1=\pm 1, \dots, \sigma_N=\pm 1} -P_2 \log_2 P_2 + \sum_{i=1}^N h_i \left( \sum_{\sigma_1=\pm 1, \dots, \sigma_N=\pm 1} \sigma_i P_2 - \langle \sigma_i \rangle \right) \\ + \sum_{1 \leq i < j \leq N} J_{ij} \left( \sum_{\sigma_1=\pm 1, \dots, \sigma_N=\pm 1} \sigma_i \sigma_j P_2 - \langle \sigma_i \sigma_j \rangle \right) + \lambda \left( \sum_{\sigma_1=\pm 1, \dots, \sigma_N=\pm 1} P_2 - 1 \right)$$

である．ここで， $P_2$  の引数は省略した．また， $h_i$ ， $J_{ij}$ ， $\lambda$  はラグランジュ未定定数である．最後の項は，確率分布の正規化に起因する制約である．式 (3.3) の  $P_2(\sigma_1, \dots, \sigma_N)$  についての偏微分を 0 と置く．これを計  $2^N$  種類ある  $P_2(\sigma_1, \dots, \sigma_N)$  のそれぞれについて行う．すると，エントロピーを最大化する分布は

$$(3.4) \quad P_2(\sigma_1, \dots, \sigma_N) = \frac{1}{Z} \exp \left( \sum_{i=1}^N h_i \sigma_i + \sum_{1 \leq i < j \leq N} J_{ij} \sigma_i \sigma_j \right)$$

と求まる． $Z$  は規格化定数であり，統計物理学で言う分配関数である．変数の勘定としては， $\lambda$  の代わりに  $Z$  を用いたと思ってよい．式 (3.4) は，統計物理学ではおなじみの，イジングモデルのボルツマン分布 (Gibbs 分布) である． $\exp$  の肩に乗っている量の符号を反転したものがエネルギーであり，いわゆる逆温度は  $h_i$ ， $J_{ij}$  に吸収されて，式 (3.4) には現れていない．

通常のラグランジュ未定定数法でそうであるように， $h_i$ ， $J_{ij}$  は制約式 (3.1)，(3.2) を満たすように決められる． $h_i$  が  $N$  個， $J_{ij}$  が  $N(N-1)/2$  個あり，この作業を解析的に行うことは難しい．ラグランジュの未定定数法の役割は，式 (3.4) の形を決定した所までで， $h_i$ ， $J_{ij}$  の値は以下のように数値的に求める． $h_i$ ， $J_{ij}$  が固定されたもとで，モデル (3.4) から計算される平均と相関は

$$(3.5) \quad \langle \sigma_i \rangle_m \equiv \sum_{\sigma_1=\pm 1} \cdots \sum_{\sigma_N=\pm 1} \sigma_i P_2(\sigma_1, \dots, \sigma_N) \quad (1 \leq i \leq N),$$

$$(3.6) \quad \langle \sigma_i \sigma_j \rangle_m \equiv \sum_{\sigma_1=\pm 1} \cdots \sum_{\sigma_N=\pm 1} \sigma_i \sigma_j P_2(\sigma_1, \dots, \sigma_N) \quad (1 \leq i < j \leq N)$$

となる．式 (3.5)，(3.6) を観測値 (3.1)，(3.2) に一致させるべく，

$$(3.7) \quad h_i^{\text{new}} = h_i^{\text{old}} + \alpha \times \text{sign}(\langle \sigma_i \rangle) \times \log \frac{\langle \sigma_i \rangle}{\langle \sigma_i \rangle_m},$$

$$(3.8) \quad J_{ij}^{\text{new}} = J_{ij}^{\text{old}} + \alpha \times \text{sign}(\langle \sigma_i \sigma_j \rangle) \times \log \frac{\langle \sigma_i \sigma_j \rangle}{\langle \sigma_i \sigma_j \rangle_m}$$

にしたがって， $h_i$ ， $J_{ij}$  を逐次的に更新する (Tang et al., 2008)． $\text{sign}$  は引数の符号， $\alpha$  は更新のステップ幅を表す．現在の  $h_i$ ， $J_{ij}$  の値が与える活動量の平均  $\langle \sigma_i \rangle_m$  (説明の便宜上，正とする) が，データが与える活動量の平均  $\langle \sigma_i \rangle$  (正とする) よりも小さければ，式 (3.7) の更新を通じて  $h_i$  が大きくなる．すると， $\langle \sigma_i \rangle_m$  が  $\langle \sigma_i \rangle$  に近づくことが期待されるのである．式 (3.8) の動作原理も同様である．

式 (3.5)，(3.6) の和の計算は  $2^N$  個の項に渡って行わなければならないので，最も計算量を要する計算過程であると言える．この部分がために，本手法を大きな  $N$  に適用することは現状では難しい．また，最急降下法よりも洗練された変数更新ルールを用いることで，それなり的高速化や精度向上を行うことはできる．ボルツマンマシンの学習の文脈で近年提案された手法の例に，Yasuda and Tanaka (2009) がある．

以上が最大エントロピー法の説明である。1次と2次の観測量 (3.1), (3.2) を用いたが, 3次以上の相関をモデルに組み入れることができる。3次相関までならば, 推定すべき変数は  $O(N^3)$  個となる。原理的には, 他の観測量を制約として追加しても, 最大エントロピー法を定式化することができる。

神経科学では, 2次より高次の相関がないと仮定したときにネットワーク全体の神経活動をどの程度表現できるのかという興味で, Schneidman et al. (2003, 2006) で最大エントロピーが最初に用いられた。約 90% の活動を式 (3.4) で説明できるという報告が複数ある (Schneidman et al., 2006; Shlens et al., 2006; Tang et al., 2008; Yeh et al., 2010)。もしデータの分布を完全に再現することによって, 実験データの平均  $\langle \sigma_i \rangle$  や相関  $\langle \sigma_i \sigma_j \rangle$  などを再現しようとする, 自明なこととして, モデルに  $2^N - 1$  個もの変数が必要となる。これが  $O(N^2)$  個ですむならば大きな倏約である。しかし, 最近の研究では  $N = 10$  程度以上に対しては2次モデルでは十分ではないことが理論的に示唆されている (Roudi et al., 2009)。データの high-order 相関や空間構造を取り入れる方法も実践されている (Santos et al., 2010; Ganmor et al., 2011)。

さて, ここまでで紹介した最大エントロピー法の大きな限界は, 時間相関を扱わないことである。実際のデータでは, 頂点  $i$  の活動が, 少し遅れて頂点  $j$  の活動を誘発するといったことが起こる。生物でも人間の社会行動でもそうであろう。最大エントロピー法をこの場合へ拡張することは, 需要があるにも関わらず, ごく近年まで扱われていなかった。最近, そのような場合を扱うための手法が Marre et al. (2009) で提案されている (Yeh et al., 2010 にも解説がある)。Marre et al. (2009) では, 既存の制約に加えて, 観測量

$$(3.9) \quad \langle \sigma_i(t) \sigma_j(t+1) \rangle \equiv \frac{1}{T-1} \sum_{t=1}^{T-1} \sigma_i(t) \sigma_j(t+1) \quad (1 \leq i, j \leq N)$$

がモデルでも再現されるという制約を課す。したがって, 仮定する分布は  $P_2(\sigma_1, \dots, \sigma_N)$  では足りず, 例えば  $P_2(\sigma_1(t), \dots, \sigma_N(t), \sigma_1(t+1), \dots, \sigma_N(t+1))$  とする必要がある, 変数が大幅に増える。解の形は, 再びボルツマン分布になり, 式 (3.9) に対応する項がエネルギー項 (式 (3.4) で述べると,  $\exp$  の肩の中身) に追加される。Marre et al. (2009) では統計物理学で言う詳細釣り合い (数学で言うと, マルコフ過程の可逆性) が成立するという強い仮定を置いて解析を行っている, 今後の手法の拡張が望まれる。

モデルが妥当で, 推定された  $J_{ij}$  が 0 から十分に離れていれば, 頂点  $i$  と頂点  $j$  はネットワーク上で隣接すると見なすことができる。単なる活動相関  $\langle \sigma_i \sigma_j \rangle$  が非零であれば頂点  $i$  と  $j$  を枝で結ぶという単純な方法も脳の研究や企業の株のネットワークの研究などで用いられることがあるが, そのような単純な方法と最大エントロピー法は異なる。よく言われるように,  $i$  と  $j$  の活動が相関していても, 例えば,  $i$  にも  $j$  にも隣接する第 3 の頂点があるからかもしれない。最大エントロピー法は, この種の影響を除去した推定手法の 1 つである。

推定された  $J_{ij}$  は, まさにネットワークの構造推定である。推定されたネットワークについて, 複雑ネットワークの分野で通常測られるような諸量を論じたり, その意義を論じたりすることはまださほど行われていないようである。その理由の 1 つに, 最大エントロピーが現状では,  $N$  が十程度までの場合にしか用いられていず, 小規模であることが挙げられる。しかし, より大きい  $N$  へと手法を拡張する試みもなされていて (Ganmor et al., 2011; Schaub and Schultz, 2012), 今後が期待される。また, 神経科学の実験を例にとって説明したが, 最大エントロピー法は, 他の様々なネットワーク上の活動観測データに適用可能であると期待される。

#### 4. おわりに

ネットワーク科学と統計学の接点という観点から, ネットワークの構造推定手法について,



静的な観測データに基づくものと、動的な観測データに基づくものについて、1つずつ手法を紹介した。

他にも、これらの解析と直接関係はしないが、会話イベントのような点過程が観測されるネットワークの研究は、近年盛んである。そのような例に、Tang et al. (2010), Isella et al. (2011), Miritello et al. (2011), Takaguchi et al. (2011)がある。日常の活動に邪魔にならないようなデバイスを身につけて参加者に一定の期間活動してもらう。すると、誰と誰がいつ、どのくらいの長さ会話したか、が多人数について、期間全体に渡ってわかるのである。そもそも、生物や、そのマクロな現象としての社会において、静的なネットワークというものはあまり存在せず、それらは極論すれば動的なネットワークの近似と見なされることが多い。しかし、近年の観測手法の発展に伴って、動的なネットワークのデータを動的な情報を保ったまま記録できるようになってきつつある。こういったデータの有効な見方はまだ定まりきってはならず、現在、世界中で研究が推進されている。

会話イベントの点過程が成すネットワークの場合も含めた上でネットワークを推定するという研究課題は、統計学とネットワーク科学の相互作用が有用でありうる課題の1つであるかもしれない。機械学習や計算機科学(や関連する統計科学)の分野で同様の問題はより古くから追求されているようである。ただ、ネットワークはしばしばコミュニティ分割されたり、そのように解釈する方が応用可能性が広がったり、頂点の次数の頻度分布がべき則に従ったり、三角形を多く持ったり、フィードフォワード構造を持ったりする。そういった実情を含めた上で統計モデルを考えることは有用かもしれない。2節で紹介したモデルは、そのような場合を有効に扱うことができるモデルの例である。

注.

<sup>1)</sup> <http://www.pnas.org/content/104/23/9564/suppl/DC1>

<sup>2)</sup> <http://www-personal.umich.edu/~mejn/>

謝 辞

原稿を読んでコメントを下さった島崎秀昭氏(理化学研究所)に御礼申し上げる。本原稿は科研費(23681033 および 20115009)の助成を受けたものである。

## 参 考 文 献

- Antal, T., Redner, S. and Sood, V. (2006). Evolutionary dynamics on degree-heterogeneous graphs, *Physical Review Letters*, **96**, 188104.
- 青山秀明, 相馬亘, 藤原義久 編(2008). ネットワーク科学への招待—世界の“つながり”を知る科学—と思考, サイエンス社(臨時別冊・数理科学 2008 年 7 月号, SGC ライブラリ 65).
- Castellano, C., Fortunato, S. and Loreto, V. (2009). Statistical physics of social dynamics, *Reviews of Modern Physics*, **81**, 591–646.
- Clauset, A., Moore, C. and Newman, M. E. J. (2008). Hierarchical structure and the prediction of missing links in networks, *Nature*, **453**, 98–101.
- Durrett, R. and Levin, S. A. (1994). Stochastic spatial models: A user's guide to ecological applications, *Philosophical Transactions of the Royal Society of London Series B*, **343**, 329–350.
- Ewens, W. J. (2010). *Mathematical Population Genetics I. Theoretical Introduction*, Springer, New York.

- Fortunato, S. (2010). Community detection in graphs, *Physics Reports*, **486**, 75–174.
- Ganmor, E., Segev, R. and Schneidman, E. (2011). Sparse low-order interaction network underlies a highly correlated and learnable neural population code, *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 9679–9684.
- Isella, L., Stehlé, J., Barrat, A., Cattuto, C., Pinton, J. P. and Van den Broeck, W. (2011). What's in a crowd? Analysis of face-to-face behavioral networks, *Journal of Theoretical Biology*, **271**, 166–180.
- Jaynes, E. T. (1957). Information theory and statistical mechanics, *Physical Review*, **106**, 620–630.
- Leicht, E. A., Clarkson, G., Shedden, K. and Newman, M. E. J. (2007). Large-scale structure of time evolving citation networks, *European Physical Journal B*, **59**, 75–83.
- Lieberman, E., Hauert, C. and Nowak, M. A. (2005). Evolutionary dynamics on graphs, *Nature*, **433**, 312–316.
- Liggett, T. M. (1985). *Interacting Particle Systems*, Springer, New York.
- Marre, O., El Boustani, S., Fregnac, Y. and Destexhe, A. (2009). Prediction of spatiotemporal patterns of neural activity from pairwise correlations, *Physical Review Letters*, **102**, 138101.
- 増田直紀 (2008a). 複雑ネットワークの研究動向について, オペレーションズ・リサーチ, **53**, 511–516.
- 増田直紀 (2008b). 複雑ネットワーク上のパーコレーション, 数学セミナー, 2008年6月～7月号(各回5頁).
- 増田直紀 (2008c). ネットワーク上の進化ゲーム, 人工知能学会誌, **23**, 652–658.
- 増田直紀 (2010). 数理で見る世の中のつながりと集まり, 数学セミナー, 2010年4月～8月号(各回2頁).
- 増田直紀, 今野紀雄 (2006). 『「複雑ネットワーク」とは何か—複雑な関係を読み解く新しいアプローチ』, 講談社, 東京.
- 増田直紀, 今野紀雄 (2010). 『複雑ネットワーク—基礎から応用まで』, 近代科学社, 東京.
- 増田直紀, 中丸麻由子 (2006). 複雑ネットワーク概説—生態学への応用を見据えて, 日本生態学会誌, **56**, 219–229.
- Masuda, N. and Ohtsuki, H. (2009). Evolutionary dynamics and fixation probabilities in directed networks, *New Journal of Physics*, **11**, 033012.
- Miritello, G., Moro, E. and Lara, R. (2011). Dynamical strength of social ties in information spreading, *Physical Review E*, **83**, 045102(R).
- Newman, M. E. J. and Girvan, M. (2004). Finding and evaluating community structure in networks, *Physical Review E*, **69**, 026113.
- Newman, M. E. J. and Leicht, E. A. (2007). Mixture models and exploratory analysis in networks, *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 9564–9569.
- Ohtsuki, H., Hauert, C., Lieberman, E. and Nowak, M. A. (2006). A simple rule for the evolution of cooperation on graphs and social networks, *Nature*, **441**, 502–505.
- Redner, S. (2001). *A Guide to First-passage Processes*, Cambridge University Press, Cambridge.
- Roudi, Y., Nirenberg, S. and Latham, P. E. (2009). Pairwise maximum entropy models for studying large biological systems: When they can work and when they can't, *PLoS Computational Biology*, **5**, e1000380.
- Santos, G. S., Gireesh, E. D., Plenz, D. and Nakahara, H. (2010). Hierarchical interaction structure of neural activities in cortical slice cultures, *Journal of Neuroscience*, **30**, 8720–8733.
- Schaub, M. T. and Schultz, S. R. (2012). The Ising decoder: Reading out the activity of large neural ensembles, *Journal of Computational Neuroscience*, **32**, 101–118.
- Schneidman, E., Still, S., Berry, M. J. and Bialek, W. (2003). Network information and connected correlations, *Physical Review Letters*, **91**, 238701.

- Schneidman, E., Berry, M. J., Segev, R. and Bialek, W. (2006). Weak pairwise correlations imply strongly correlated network states in a neural population, *Nature*, **440**, 1007–1012.
- 島崎秀昭 (2011). 対数線形モデルによるマルチニューロンスパイクデータ解析, 日本神経回路学会誌, **18**, 190–199.
- Shlens, J., Field, G. D., Gauthier, J. L., Grivich, M. I., Petrusca, D., Sher, A., Litke, A. M. and Chichilnisky, E. J. (2006). The structure of multi-neuron firing patterns in primate retina, *Journal of Neuroscience*, **26**, 8254–8266.
- Sood, V., Antal, T. and Redner, S. (2008). Voter models on heterogeneous networks, *Physical Review E*, **77**, 041121.
- Szabó, G. and Fáth, G. (2007). Evolutionary games on graphs, *Physics Reports*, **446**, 97–216.
- Takaguchi, T., Nakamura, M., Sato, N., Yano, K. and Masuda, N. (2011). Predictability of conversation partners, *Physical Review X*, **1**, 011008.
- 田中和之 (2006). 『確率モデルによる画像処理技術入門』, 森北出版, 東京.
- Tang, A., Jackson, D., Hobbs, J., Chen, W., Smith, J. L., Patel, H., Prieto, A., Petrusca, D., Grivich, M. I., Sher, A., Hottowy, P., Dabrowski, W., Litke, A. M. and Beggs, J. M. (2008). A maximum entropy model applied to spatial and temporal correlations from cortical networks in vitro, *Journal of Neuroscience*, **28**, 505–518.
- Tang, J., Scellato, S., Musolesi, M., Mascolo, C. and Latora, V. (2010). Small-world behavior in time-varying graphs, *Physical Review E*, **81**, 055101(R).
- Yasuda, M. and Tanaka, K. (2009). Approximate learning algorithm in boltzmann machines, *Neural Computation*, **21**, 3130–3178.
- Yeh, F. C., Tang, A., Hobbs, J. P., Hottowy, P., Dabrowski, W., Sher, A., Litke, A. and Beggs, J. M. (2010). Maximum entropy approaches to living neural networks, *Entropy*, **12**, 89–106.

## Statistical Methods for Inferring Network Structure

Naoki Masuda

Department of Mathematical Informatics,  
Graduate School of Information Science and Technology, The University of Tokyo;  
PRESTO, Japan Science and Technology Agency

There have been many researches on networks during the past 15 years. This research field is referred to as network science or complex networks research. The term “network” in this context is equivalent to the term “graph” in mathematical graph theory. Data of many actual graphs are used in various fields. Independently of social network analysis researchers, researchers in statistical physics, applied mathematics, and web engineering, in particular, have started studying networks. Now, people from different fields not limited to the abovementioned fields are engaged in analysis of networks. Because network science deals with real data of networks, statistical sciences will find various applications in network researches. This review paper briefly surveys two maximum likelihood methods for inferring network structures from data. First, it explains the maximum likelihood method proposed by Newman and Leicht. They assumed that the nodes in an observed graph are partitioned into groups, and nodes in the same group tend to have similar connectivity to other nodes. Then, they established an EM algorithm to estimate the partition of the nodes and the parameter values that determine the likelihood with which nodes in certain groups are connected to each other. Second, it briefly introduces the maximum likelihood method based on a maximum entropy model. Although it is a classical approach, the method has been applied to analysis of neural activity data.